

DOCUMENT RESUME

ED 123 255

TM 005 314

AUTHOR Smilansky, Jonathan; Guerin, Robert O.
TITLE Minimal Acceptable Performance Levels for Criterion Referenced Multiple Choice Examinations and Their Validation.
PUB DATE [Apr 76]
NOTE 6p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
DESCRIPTORS Complexity Level; Criteria; *Criterion Referenced Tests; *Cutting Scores; Measurement Techniques; *Multiple Choice Tests; Response Style (Tests); Statistical Analysis; *Test Construction; *Test Validity

ABSTRACT

A method for establishing minimal acceptable performance levels for criterion referenced multiple choice examinations is described and two procedures for evaluating its validity are discussed. The method involves assigning a priori item difficulties based on the similarity between the distractor and the correct response. In estimating the validity of this method it was found that a priori assigned distractor weights were significantly related to the choice patterns of different levels of students. In a second analysis it was found that the a priori set item difficulties were significantly related to normative data regarding item difficulty. (Author/DEP)

* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *

ED123255

ABSTRACT

Minimal Acceptable Performance Levels For Criterion Referenced Multiple Choice Examinations and Their Validation

Jonathan Smilansky, M.A.
Evaluator of Training Programs
Center for Educational Development
College of Medicine, University of Illinois

Robert O. Guerin, Ph.D.
National Board of Medical Examiners
Philadelphia, Pennsylvania

This paper describes a method for establishing minimal acceptable performance levels for criterion referenced multiple choice examinations and two procedures for evaluating its validity.

The method involves assigning a priori item difficulties based on the similarity between the distractor and the correct response.

In estimating the validity of this method it was found that the a priori assigned distractor weights were significantly related to the choice patterns of different levels of students. In a second analysis it was found that the a priori set item difficulties were significantly related to normative data regarding item difficulty.

Session No. D - 20: Criterion Referenced Testing

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

TM005 314

Minimal Acceptable Performance Levels
For Criterion Referenced Multiple
Choice Examinations and Their
Validation

This paper describes a method for establishing minimal acceptable performance levels for criterion referenced multiple choice examinations and two procedures for evaluating the validity of this method.

One of the important issues in designing a criterion referenced examination is the decision regarding the passing score, that is, the minimal acceptable level of performance. With most criterion referenced examinations, a general rule of thumb is usually evoked, such as 80% or 90% success, to determine a passing score. This approach does not take into consideration the difficulty of the items that make up a specific test. One possible way to consider item difficulty is by judging the similarity between the distractors and the correct response. As the distractors get more similar it becomes harder to distinguish them from the correct response thereby making the item more difficult. Item difficulty, in this sense, is criterion referenced since it relates to content rather than to its performance relative to a specific group of students. The method used to determine item difficulty, as well as two procedures used to compare these a priori item difficulty estimates are described in this paper.

The faculty of the School of Basic Medical Sciences at the University of Illinois Chicago Campus (SBMS-MC), prepare multiple choice items for diagnostic examinations administered at the end of the fall and winter quarters of the freshman year. A certification examination prepared by both the Chicago and Champaign-Urbana faculties is given at the end of the first year. Examination topics are generally organized according to both organ systems and disciplines such as anatomy, biochemistry, genetics, histology, and physiology. A blueprint is developed specifying the proportion of items in each topic or discipline according to the relative importance of each. About twenty-five authors contribute items to a quarterly examination. The results of the two quarterly examinations are used to provide the faculty with evidence of student progress and the student with diagnostic feedback on their achievement. The results of the certification examination are used by promotion committees as one criterion for advancement from the basic science to the clinical portions of the medical school curriculum. An external a priori criterion, the Minimal Acceptable Performance (MAP), is established by the SBMS faculty to differentiate barely passing from failing students, without further discrimination in terms of degree of achievement.

The model proposed by Nedelsky (2) to determine an external (absolute) minimal acceptable performance level has been adopted by the college faculty (3). One assumption of this model is that students may be classified into clearly passing, barely passing and clearly failing groups; a second is that faculty can define these levels of achievement by assigning difficulty codes to all distractors of each item prior to the administration of the examination. An option which a minimally passing student should recognize as incorrect is coded zero; one which a minimally passing student should not recognize as incorrect is coded two; all other options are coded one. The correct response is also assigned a code of two. The Minimum Pass Index (MPI) of an item is the probability that a minimally passing student would select the correct option by chance. Quantitatively, the MPI is the sum of the difficulty codes divided into two.

Second Method

For each of the diagnostic examinations and for the certifying examination, a correlation was computed between the a priori established Minimal Pass Indices for each item and the item difficulty.

Results - First Method

Table 1 summarizes the findings on each of the three tests. In all there is a significant positive correlation between total score and the selection of code 2 options ("educated mistakes"). There is also a negative correlation between total score and the proportion of code 0 options selected. The frequency of code 1 options was not related to total score in any test. In general, these correlations support the hypothesis that the more successful students make more "educated mistakes", that is, they choose more frequently those distractors judged by the item authors to be more difficult to discriminate from the correct response.

INSERT TABLE 1 ABOUT HERE

Results - Second Method

Table 2 summarizes the results on each of the three examinations. In all there is a significant positive correlation between the a priori set Minimal Pass Indices and the item difficulties. It is also interesting to note the similarity in means, standard deviations, Minimal Pass Indices, item difficulties and correlations for all three examinations given the large number of authors writing items for these examinations.

INSERT TABLE 2 ABOUT HERE

Summary and Conclusions

This paper describes a method for establishing minimal acceptable performance levels for criterion referenced multiple choice examinations and two different systems for evaluating the validity of this method. The first method examined the accuracy of distractor weights by comparing them to choice patterns of different levels of students. The second method compared the a priori set Minimal Pass Indices for each item to normative data regarding item difficulty.

Results of the first analysis suggests that as student achievement increases so does their tendency to make more "educated mistakes", i.e., a shift from choosing distractors labeled clearly incorrect to choosing those labeled more nearly correct. Results of the second analysis suggest positive correlation between the a priori determined Minimum Pass Index of an item and the empirical item difficulty. Such consistency seems particularly impressive considering the large number of faculty members who write items for each examination. These results enhance confidence in the validity of the methods described to compute predetermined passing standards for criterion referenced examinations.

References

1. Bobula, J. Minimum Pass Level Study. Report to the Faculty, Center for Educational Development, College of Medicine, University of Illinois, 25-31, 1974.
2. Nedelsky, L. Absolute Grading Standards for Objective Tests. Educational and Psychological Measurement, 14: 3-19, 1954.
3. Setting Standards of Competence: The Minimum Pass Level, Committee on Student Appraisal, University of Illinois, College of Medicine, January, 1964.
4. Taylor, D.D. and Reid, J.C. Criterion-Referenced Evaluation of Student Performance, Journal of Medical Education, 47: 970-971, 1972.

TABLE 1

Means, Standard Deviations and Product Moment Correlations for the December 1973, March 1974 and June 1974 Examinations

Variable		Mean	S.D.	Correlation With Total Score
INCORRECT RESPONSES	0 Wgt. Dec. '73	.28	.06	-.28*
	Mar. '74	.23	.06	-.35*
	June '74	.31	.03	-.31*
	1 Wgt. Dec. '73	.40	.06	-.10
	Mar. '74	.44	.06	-.11
	June '74	.44	.03	.09
	2 Wgt. Dec. '73	.27	.05	.50*
	Mar. '74	.31	.06	.49*
	June '74	.22	.03	.33*
CORRECT RESPONSES		Dec. '73	.10	
		Mar. '74	.11	
		June '74	.09	

* Significant at $p < .01$

TABLE 2

Means, Standard Deviations and Correlations Between A-Priori Set Minimal Passing Indices and Item Difficulties for the December 1973, March 1974 and June 1974 Examinations

	Minimal Pass Indices		Item Difficulties		Correlation between MPI and Item Difficulty
	Mean	S.D.	Mean	S.D.	
December '73 (189 items)	.44	.15	.65	.21	.33*
March '74 (168 items)	.40	.12	.63	.21	.29*
June '74 (452 items)	.43	.13	.61	.21	.22*

*Significant at $p < .01$